Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-34-RC3, 2019

Atmospheric Measurement Techniques Discussions

Interactive comment on "Calibration of the 2007–2017 record of ARM Cloud Radar Observations using CloudSat" *by* Pavlos Kollias et al.

Roger Marchand (Referee)

rojmarch@uw.edu

Received and published: 5 May 2019

Review of Manuscript amt-2019-34

Title: Calibration of the 2007-2017 record of ARM Cloud Radar Observations using CloudSat

Authors: Kollias, Treserras and Protat

Overview:

I think publication of this technique and the results for ARM radars will be of value to many investigators and investigations that have (and will continue to) rely on ground-

based radar datasets. While I was under no illusion that the ARM radars were well calibrated, I nonetheless found the results to be sobering.

Recommendation: Publish after minor revisions

General Comments:

1) A few more details on the technique.

I largely follow the technique, but you need to add a few more details, see specific comments for Page 6. The goal should be to make it so that someone else could implement the approach given this description. In particular, please discuss uncertainties in estimated calibration corrections associated with equations 2 and 3, as well as the height range used to estimate the best offset.

2) Differences in rules and thresholds for including or not including columns & Verification metrics.

Do the different rules and thresholds for including or not including radar columns, which are to some degree necessarily different between CloudSat and ARM, matter? (see e.g. differences on Page 5, line 16; Page 7 line 17). I am concerned about the possibility that differences in the mean Z-profiles might be due simply to having different conditions or "distributions of cloud-types" in each collection from which the mean-Z-profile is calculated.

One way to check this would be to look not just at the mean-Z-profile but also to ask if the two profiles are based on a similar fraction of the observations in each set. Said another way, once you construct your "non-precipitating CFAD" and pick your dBZe-threshold (for calculating the mean-Z-profile), is the profile of cloud fraction associated with this dBZ-threshold the same for both CloudSat and ARM. If it is, then one can be confident that errors in the reflectivity correction due to differences in cloud populations will be small.

I suggest creating a metric, such as the vertically integrated absolute cloud-fraction

difference divided by the mean cloud fraction, and plotting this information along with the calibration corrections. Likewise, it would be interesting to see how this metric depends on the number of columns, which like figure #6, should give one a sense of what is a reasonable value for this quantity.

Likewise, the cloud-top-height (CTH) distribution comparison you introduce (Fig 4d) provides confidence that the calibration correction is robust and that it is based on the same cloud populations. As far as I can see, after you introduce the idea of this as "a verification", you don't use it. At a minimum it seems like you should discuss whether the CTH distributions are consistently improved (made more similar) with the radar correction or not. Again, you might make a metric that expresses this improvement – though I suspect the above cloud fraction metric is likely better for this purpose.

3) Results for Darwin and the size of the analysis region

I don't typically like to point to my own work when reviewing an article, but in this case I think some work that a former student of mine Zheng Liu, Tom Ackerman and I have done at Darwin is very germane to this study.

Liu, Z., R. Marchand, and T. Ackerman (2010), A comparison of observations in the tropical western Pacific from groundâĂŘbased and satellite millimeterâĂŘwavelength cloud radars, J. Geophys. Res., 115, D24206, doi:10.1029/2009JD013575.

In particular, we compared CloudSat and the Darwin MMCR measurements and we investigated the size of the analysis region and sampling uncertainties in some detail. That study very much supports using a 300 km radius area and 6 month window.

Note also figure 8 in this paper. While we did not derived a calibration offset, our results are broadly consistent with idea that ARM calibration was too LOW at Darwin in the 2006-7 (wet season), and agrees your with figure 9a (CloudSat – ARM difference of $\sim$= 5 dB at this time).

Minor Comments:

Page 1, line 28. detail => detailed

Page 2, line 5. Opening sentence is awkward, rephrase. Perhaps "Part of the motivation for the ARM radar expansion, was to improve cloud microphysical retrievals through the use of dual-wavelength ratios, that is, making use of the relative difference in radar scattering at different wavelengths. This difference signal is often only a few dB and as one might expect, this .... "

Page 2, line 11. Do you mean calibrating ARM vertically pointing radars is more difficult than the WSR-88D network? What is being compared to what is not clear? Suggest you rephrase this and following sentence to be clearer and generally read better.

Page 3, line 3. Change ". . . is such diverse" to ". . . to such a diverse set . . .".

Page 3, line 31. ". . .on the same. . ." to "at the same".

** Page 4: First paragraph: Mode analysis.

How did you account for differences in the minimum detectable signal between the modes? I presume you only included neighboring time & range bins where both modes have a measurement with high SNR?

Page 4. The first paragraph launches into a discussion of mode differences (which is useful) but a bit confusing when one is expecting a comparison of CloudSat and ARM calibrations. I suggest breaking this paragraph about line 8 and adding.

"Therefore as a prelude to comparing CloudSat and ARM, we begin with a comparison of reflectivity values between ARM radar modes. As will become clear later, changes in the reflectivities between modes is often, though not always, indicative of changes in overall calibration."

Page 4, line 13. I presume "bid" should be "big". Perhaps

Page 4, line 14. Specify period (6 months?).

Page 4, line 16. Perhaps change to read: “Overall, the mode reflectivity differences are small ($\pm$2 dB) and only occasionally are the differences much higher than 2 dB. While the absolute values of mode difference is in the next generation of ARM cloud profiling radars (KAZR and KAZR2) is often similar, arguably there are fewer jumps or rapid changes (except perhaps at OLI). In general it is difficult to identify which mode has a better calibration, because as will be shown, the calibration difference between CloudSat and ARM is typically larger than $\pm$2 dB.

** Page 4: Second paragraph on difference between Protat and current approach.

This paragraph is nearly impossible to follow if you don’t already know what Protat 2010 did. In particular, I have no idea what “. . . a rigorous selection of the CloudSat overpasses within a certain radius to avoid any errors in the estimation of the proximity of CloudSat columns to the ARM site location” means. But other parts of this are confusing to me (and I am familiar with Protat 2010). I strongly encourage you to reorganize the manuscript such that you FIRST explain your approach in detail and ONLY at the end of this material highlight how this approach differs from Protat (2010).

** Page 6, equation 2. What is the justification for using a constant here? You used Rosenkranz and . . . ? Somehow you must have specified some set of atmospheric profiles to come up with these constants? Explain in enough detail so someone else could implement this idea. Nominally, I think it would have been better to calculate a set of gasses corrections for ARM (perhaps using ERA data just as CloudSat does). **But in lieu of this, I think you need to address how much error (uncertainty) using this constant introduces in your calibration correction.

** Page 6, equation 3. What assumptions does this equation entail? Again, I think you need to address how this impacts the uncertainty of your calibration corrections?

** Page 6, line 17. I am not sure I understand this definition, there seems to be a grammatical error here. Do you simply mean "precipitaing column = 10% (or more) of the radar volumes below the FL have ANY reflectivity (even if it is -30 dBZe). So if FL

is 2 km at there are 20 bins below 2km, if 3 bins have ANY detection (even if it is just a low cloud) you are calling it precipitating?

** What do you do if the FL is near the surface (in the CloudSat clutter) or with only a few ARM bins $\sim$ 5 ?

** Why do you use a different threshold 35% for CloudSat? This seems arbitrary.

** Page 6, items 2 to 4. This material seems important and needs to be better explained. In particular, how does the "degradation" work? I presume you mean that the mean-Z-profiles, are obtained from the CFADS in step 4 by summing bins with dBZ > Threshold > Minimum Detectable Signal (MDS) weighted by the bin dBZe? (I note without weighting this just give you the “profile of cloud fraction”). If yes, it might be important to choose threshold that is + 3 to 5 dB larger than the MDS.

** Where/How does the SNR > -15 mentioned early come in?

**Page 7, item 2. My experience at Darwin suggests (and your example in Fig. 4) that the height range used might matter here. How much does the estimated correction change in this example if you change the range form 3 to 12 km, that is, 10 to 12 Km and 3 to 4 km?

Page 7, line 13. What does "maximum fraction of ... warm temp" mean ? I don’t follow.

Page 7. It seems you address the issues of the number of columns in detail later in the text, but do NOT the distance issue. (see also general comment #3). Perhaps add some discussion and/or better yet show result for OLI site (where you have lots of data) – add a line to fig 7 – for results based on 100 vs. 300 km?

** Page 7/8, analysis on number of columns vs. number of good columns ?

I like very much the analysis you have included on the number of columns. But unless I misunderstand you are counting ALL columns here. Not the number of good columns (i.e. columns which are devoid of high/ice clouds or precipitating). I think it would be

far more sensible to count the number of columns with good data (and set a minimum threshold on this) rather than all radar columns.

Page 9, line 21. I presume you mean “only” not “on”. Why is it that only GE mode is available?

Page 9, line 22. The stability here demonstrates that changes in mode differences are “indicative” not necessary for there to be calibration issues.

Page 10, line 37. So the dots here in Fig. 11 represent different frequencies, not just different months? I strongly suggest using different symbols for the different frequencies.

Page 11, line 27. Perhaps rephrase as "In many cases, the offset ... . Thus, changes in the reflectivity offset between the modes should be monitored, and used to identify periods where the calibration stability is suspect, and moving forward perhaps trigger more prompt additional external calibration evaluations".

Page 12, line 4. Seem redundant with the above comments on page 11.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-34, 2019.